Unified Minimax Optimization Framework for Propensity Score Estimation in Debiased Recommendation

Chunyuan Zheng[1], Haocheng Yang[2], Jinkun Chen[3], Shufeng Zhang[4], Tianyu Xia[1*]

[1]Peking University, Beijing, China
[2]National University of Singapore, Singapore
[3]Dalhousie University, Halifax, Canada
[4]University of North Carolina at Chapel Hill, North Carolina, USA
cyzheng@stu.pku.edu.cn, 2311110185@bjmu.edu.cn

Abstract

Recommendation systems commonly face selection bias from missing-not-at-random (MNAR) collected data. To address this bias, propensity-based methods such as inverse propensity scoring (IPS) and doubly robust (DR) estimators are widely used. In addition, many methods extend the vanilla IPS and DR to further control the bias, variance, propensity mis-calibration, and imbalance, but they only optimize some of the above metrics, limiting the debiasing performance. In this paper, we first empirically find that controlling one metric cannot guarantee the control of other important metrics, then we reveal a fundamental structural commonality among the above four important metrics, and propose a Unified Propensity Optimization (UPO) framework that optimizes all metrics simultaneously by a minimax optimization algorithm. Theoretically, we demonstrate that minimizing the UPO loss effectively controls all metrics, ensuring their simultaneous improvements without incurring additional bias, and achieving reduced variance compared to naively adding up multiple control losses in penalty terms. Empirically, experiments on a semi-synthetic dataset and three real-world datasets validate UPO's effectiveness, demonstrating superior performance compared to state-of-the-art methods with minor computational overhead. We fully open-source our code.

Code — https://github.com/yhc-666/UPO

Introduction

Recommendation systems (RS) are integral to personalized decision-making in domains such as e-commerce, content platforms, and healthcare (Yang et al. 2018; Zheng et al. 2022; Huang et al. 2023; Su et al. 2023; Lin et al. 2025a,b). A critical issue affecting the performance of RS is selection bias arising from missing-not-at-random (MNAR) in collected training data, due to users are free to choose items to provide feedback, resulting in the training data not a representative of the target data (Ai et al. 2018; Wang et al. 2020; Zhang et al. 2024; Zheng et al. 2025b; Zhang et al. 2025).

Causal Inference is widely used to address the MNAR problem (Li et al. 2023e; Zhou et al. 2025b; Wu et al. 2025; Yang et al. 2025; Huang et al. 2025a,b). Specifically, Inverse propensity scoring (IPS) addresses MNAR by reweighting observed samples using learned propensity scores (Schnabel et al. 2016; Saito et al. 2020). Doubly robust (DR) combines propensity reweighting with error imputation, which is unbiased if either the propensity score or the imputed error is correct for all user-item pairs (Wang et al. 2019b). Nevertheless, IPS and DR estimators still struggle with the following four aspects: high bias, high variance, mis-calibration, and inadequate feature balancing (Imai and Ratkovic 2014; Guo et al. 2017; Bonner and Vasile 2018; Kweon and Yu 2022).

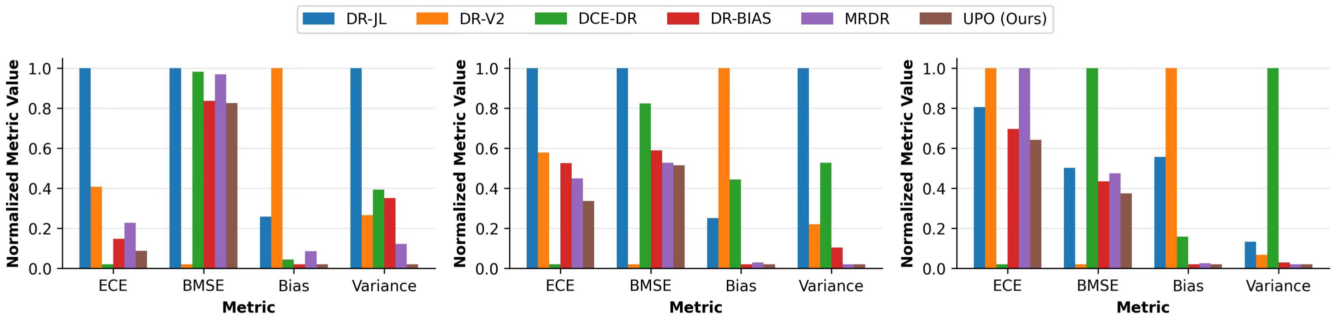
Many DR-based methods extend the vanilla DR to mitigate the above issue, but they only control some of the metrics: DR-BIAS minimizes bias by adding bias as a constraint term in DR loss (Dai et al. 2022), MRDR reduces variance by adopting variance as an imputation model loss (Guo et al. 2021), and DR-Var, DR-GPL, and UMVUE-DR reduce bias and variance simultaneously by adding both constraints in DR loss (Dai et al. 2022; Zhou et al. 2023; Zheng et al. 2024). In addition, DCE-DR improves calibration by proposing a mixture of expert structures in propensity and imputation models (Kweon and Yu 2024), and DR-V2 promotes feature balancing by minimizing balanced mean squared error between observed and unobserved samples under manually selected balancing functions (Li et al. 2023d).

However, all four metrics are important, and we claim that control some of them cannot guarantee the performance of remaining. To demonstrate this issue, we conduct a semi-synthetic experiment on the MovieLens-100K (ML-100k) dataset (Harper and Konstan 2015). We impute missing ratings and generate true propensity scores (details in Experiments Section), then evaluate six methods: four DR variants optimizing specific metrics, one baseline DR-JL method, and our proposed UPO on four metrics. Results in Figure 1 show that optimizing one metric cannot guarantee the performance of remaining. For instance, DCE-DR achieves favorable calibration but underperforms in other metrics.

To address these limitations, we propose a Unified Propensity Optimization (UPO) framework that optimizes all metrics simultaneously by a minimax optimization algorithm. Specifically, our contributions include:

- We reveal a structural commonality among four critical

[*]Corresponding author.



(a) $p = 0.4$

(b) $p = 0.5$

(c) $p = 0.6$

Figure 1: The performance of four metrics and six DR-based debiasing estimators under varying simulated propensity. The propensities are simulated by p^k, where $k = 1, 2, 3, 4, 4$ corresponds to ratings $\{5, 4, 3, 2, 1\}$. The methods from left to right are DR-JL (blue), DR-V2 (orange), DCR-DR (green), DR-BIAS (red), MRDR (purple), and UPO (ours, brown).

metrics (bias, variance, calibration, and balancing), then propose a unified minimax optimization algorithm.

- Theoretically, we demonstrate that minimizing the UPO loss effectively controls all metrics, ensuring their simultaneous improvements without incurring additional bias, and achieving reduced variance compared to naively adding up multiple control losses in penalty terms.
- Empirically, experiments on a semi-synthetic dataset and three real-world datasets validate UPO's effectiveness, demonstrating superior performance compared to state-of-the-art methods with minor computational overhead.

Related Work

Selection Bias. Selection bias arises when user feedback is MNAR, making observed interactions systematically different from the full preference distribution (Chen et al. 2022; Li et al. 2024b,d; Pan et al. 2025). Inverse Propensity Scoring (IPS) method addresses this issue via reweighting each observed sample reciprocally by its probability of being observed, but it suffers from instability and high variance given extreme propensities (Schnabel et al. 2016). Doubly robust (DR) method integrate IPS and error imputation, but it still retains issues of large variance, propensity mis-calibration, and insufficient feature balancing (Wang et al. 2019b; Saito 2020b). Beyond classical IPS/DR estimators, the community has explored several directions. Weak-/semi-supervised learning and counterfactual modeling approaches aim to leverage unlabeled feedback (Wang et al. 2025b; Zhou et al. 2025a; Wu et al. 2025). Recent works focus specifically on propensity estimation refinement, a shared component in IPS/DR (Ma et al. 2018; Ding et al. 2022; Wang et al. 2022; Li et al. 2023a; Wang 2024; Li et al. 2024c; Wang et al. 2024; Zheng et al. 2025a). One of its research lines devotes to directly regularizing key evaluation metrics of learned propensities (Zhou et al. 2025a; Wang et al. 2025a).

Feature Balancing. A well-estimated propensity model should balance features distributions across observed and unobserved samples (Imai and Ratkovic 2014). IPS-V2/DR-V2 enforce this balance via Balanced Mean Squared Error (BMSE) metric (Li et al. 2023d, 2024a). Its kernel-based extensions (AKBIPS/AKBDR) adaptively optimize balancing in reproducing kernel Hilbert spaces, controlling both bias and variance (Li et al. 2024e).

Calibration. Calibration concerns how well predicted propensities reflect true probabilities of being observed (Wang et al. 2025c). Foundational work such as temperature scaling and Bayesian binning provides a standard evaluation metric as expected calibration error (ECE) (Naeini, Cooper, and Hauskrecht 2015; Guo et al. 2017). Recent RS-specific approaches include DCE-DR, which improves calibration through a mixture-of-experts structure and evaluate performance by the bin-wise ECE (Kweon and Yu 2024); Cali-MR, which employs gradient-based bi-level calibration optimization (Gong and Ma 2025); and uncertainty-based frameworks tailored for conversion-rate prediction tasks (Hu et al. 2025).

Bias-Variance and Stability. Estimators face intrinsic bias-variance trade-offs. SNIPS, MRDR, DR-BIAS and DR-MSE optimize for variance, bias, or joint bias-variance criteria (Swaminathan and Joachims 2015; Guo et al. 2021; Dai et al. 2022). MR and StableDR ensure stability under severe MNAR, while TDR-JL optimizes robustness for collaborative filtering (Li et al. 2023b; Li, Zheng, and Wu 2023; Li et al. 2023c). Recent hybrid objectives jointly optimize multiple criteria, like bias-variance (MRDR-GPL) and calibration-balancing (CBPL) (Zhou et al. 2023; Zhang and Xia 2025). However, existing approaches optimize criteria in isolation, and no prior work jointly integrates bias, variance, calibration, and feature balancing, motivating a unified minimax propensity optimization framework.

Preliminary

Problem Setup

We adopt the potential outcome framework to formally characterize selection bias in recommender systems. Let $\mathcal{U}$ and $\mathcal{I}$ denote user and item sets, respectively. The target user-item interaction space is $\mathcal{D} = \{(u, i) | u \in \mathcal{U}, i \in \mathcal{I}\}$. For each pair $(u, i) \in \mathcal{D}$, $x_{u,i} \in \mathbb{R}^d$ represents user-item pair feature. $o_{u,i} \in \{0,1\}$ is a binary indicator to indicate if user u rates item i. If $o_{u,i}=1$, we can observe the corresponding rating $r_{u,i}$, otherwise, the $r_{u,i}$ is missing. Our goal is to predict ratings for all pairs. Ideally, if we can observe all ratings, the prediction model $\hat{r}_{u,i}=f(x_{u,i};\theta)$ can be trained by minimizing the following ideal loss:

$$\mathcal{L}_{\text{ideal}}(\theta)=\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}} e_{u,i},$$

where $e_{u,i}=\ell(f(x_{u,i};\theta),r_{u,i})$ is the prediction loss. In practice, ratings are partially observed with non-random missingness. Minimizing loss only over observed samples $\mathcal{O}=\{(u,i)|(u,i)\in\mathcal{D},o_{u,i}=1\}$ yields suboptimal predictions, due to such a loss function is a biased estimation of the ideal loss under data MNAR.

Propensity-Based Methods

To address selection bias, propensity-based methods are widely adopted. Specifically, the propensity score is defined as $p_{u,i}=\mathbb{P}(o_{u,i}=1\mid x_{u,i})$, i.e., the probabilities of observing ratings. We define $\hat{p}_{u,i}$ as the learned propensity score via a propensity model $h_\psi(x)$. The Inverse Propensity Score (IPS) method uses propensity to reweight observed samples, training the prediction model by minimizing:

$$\mathcal{L}_{\text{IPS}}(\theta)=\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\frac{o_{u,i}e_{u,i}}{\hat{p}_{u,i}},$$

which is unbiased when $\hat{p}_{u,i}=p_{u,i}$. Doubly robust (DR) method further introduces an error imputation model, training the prediction model by minimizing:

$$\mathcal{L}_{\text{DR}}(\theta)=\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[\hat{e}_{u,i}+\frac{o_{u,i}(e_{u,i}-\hat{e}_{u,i})}{\hat{p}_{u,i}}\right], \quad (1)$$

where $\hat{e}_{u,i}=g_\phi(x_{u,i})$ is the imputation of the prediction error $e_{u,i}$, learned by the following imputation loss

$$\mathcal{L}_{\text{imp}}(\phi)=\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left[\frac{o_{u,i}(e_{u,i}-\hat{e}_{u,i})^2}{\hat{p}_{u,i}}\right]. \quad (2)$$

DR-based estimators remain unbiased if either $\hat{p}_{u,i}=p_{u,i}$ or $\hat{e}_{u,i}=e_{u,i}$. However, both methods heavily depend on accurate propensity learning and can suffer from high bias, high variance, mis-calibration, and feature imbalance if propensities are inaccurate. Due to IPS is a special case of DR with $\hat{e}_{u,i}=0$, we only discuss DR in the following.

Methodology

Recall that we empirically show control of one of the metrics cannot guarantee the remaining in Figure 1. Therefore, we propose a unified minimax optimization framework, termed Unified Propensity Optimization (UPO), to make the learned propensity simultaneously lead to small bias, small variance, well-calibration, with covariance balancing property. Specifically, we first introduce the definition of four metrics and reveal the structural commonality among them, then we introduce the proposed minimax algorithm in detail with theoretical results.

Definition and Commonality of Four Metrics

- **Bias:** The bias of DR is defined as (Wang et al. 2019a):

$$\mathcal{L}_{\text{Bias}}=\frac{1}{|\mathcal{D}|}\left|\sum_{(u,i)\in\mathcal{D}}\frac{(o_{u,i}-\hat{p}_{u,i})(e_{u,i}-\hat{e}_{u,i})}{\hat{p}_{u,i}}\right|.$$

- **Variance:** The variance of DR is (Guo et al. 2021):

$$\mathcal{L}_{\text{Var}}=\frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\frac{p_{u,i}\left(1-p_{u,i}\right)}{\hat{p}_{u,i}^2}\left(e_{u,i}-\hat{e}_{u,i}\right)^2. \quad (3)$$

- **Expected Calibration Error (ECE):** The ECE of propensity model is defined as (Kweon and Yu 2024):

$$\mathcal{L}_{\text{ECE}}=\sum_{m=1}^{M}\frac{|B_m|}{|D|}\left|\frac{\sum_{(u,i)\in B_m}o_{u,i}}{|B_m|}-\frac{\sum_{(u,i)\in B_m}\hat{p}_{u,i}}{|B_m|}\right|,$$

  where M is the number of bins and B_m is the disjoint split bins. For example, if we use the equal-width to split bin, then $B_m=\left\{\hat{p}_{u,i}:\hat{p}_{u,i}\in\left(\frac{m-1}{M},\frac{m}{M}\right]\right\}$. This metric measures the degree of mis-calibration in each pre-specified bin.

- **Balanced Mean Squared Error (BMSE):** The BMSE of propensity model in DR is defined as (Li et al. 2023d):

$$\mathcal{L}_{\text{BMSE}}=\left\|\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left(\frac{o_{u,i}}{\hat{p}_{u,i}}-\frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right)\phi(x_{u,i})\right\|^2,$$

  where $\phi(x)$ is an arbitrary feature balancing function. This metric measure the degree of feature balance of the propensity model.

To control the above metrics, a naive way is to regard the four loss functions as constraints during propensity learning. However, the weight of each constraint is hard to determine. In addition, due to there is infinite bin-splitting ways and infinite choices of $\phi(x)$, we cannot calibration every possible situation. Therefore, to control all of them, we need first to find the commonality between each metric. Note that Bias, ECE, and BMSE include $o_{u,i}-\hat{p}_{u,i}$, such a similar structure motivates us to integrate these metrics into a single loss function, and to adopt a unified minimax optimization framework to minimize the worst case adaptively. Before introducing the framework, we need one more step for Variance approximation, to transform the original Variance to such similar structure[1].

Lemma 1 (Variance Approximation) *After taking the main component in Taylor expansion of variance, $\mathcal{L}_{Var}$ defined in equation* (3)*, around the true propensity score p, we can obtain the following loss function for propensity model training:*

$$\mathcal{L}_{\textit{Var-Tay}}=2\frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\frac{1-p_{u,i}}{p_{u,i}^2}(o_{u,i}-\hat{p}_{u,i})(e_{u,i}-\hat{e}_{u,i})^2.$$

[1]All proofs can be found in the Appendix in arXiv version.

However, due to we cannot assess the $p_{u,i}$ in practice, following (Guo et al. 2021), we propose the following unbiased empirical form for propensity learning when $\hat{p}_{u,i} = p_{u,i}$:

$$\mathcal{L}_{\text{Var-Emp}} = 2\frac{1}{|\mathcal{D}|^2}\sum_{(u,i)\in\mathcal{D}}\frac{1-\hat{p}_{u,i}}{\hat{p}_{u,i}^2}(o_{u,i}-\hat{p}_{u,i})(e_{u,i}-\hat{e}_{u,i})^2.$$

The Unified Minimax Optimization Framework

Due to the calibration metrics require a bin splitting, we first split the propensity range $[0, 1]$ into M bins by pre-defining the $M-1$ split point. We can use equal-width split to make the width of each bin as $1/M$, or we can use equal-mass split to ensure the same sample size in each bin. Then, we propose the UPO loss as below:

$$\mathcal{L}_{\text{UPO}} = \frac{1}{|\mathcal{D}|}\sum_{m=1}^{M}\left|\sum_{(u,i)\in\mathcal{B}_m} w_{u,i}(o_{u,i}-\hat{p}_{u,i})\right|, \quad (4)$$

where $w_{u,i}$ are trainable weights parameterized by a neural network. The Lemma below shows the relationship between the proposed $\mathcal{L}_{\text{UPO}}$ and each metric.

Lemma 2 (Upper-Bound Relationship) *For $\mathcal{L}_{upo}$ defined as Equation (4), the following upper-bound relationship hold true by Jensen's inequality:*

$$\mathcal{L}_{upo} = \mathcal{L}_{ECE} \quad \text{for } w_{u,i} = 1,$$
$$\mathcal{L}_{upo} \geq \sqrt{\mathcal{L}_{BMSE\text{-}abs}} \quad \text{for } w_{u,i} = \frac{\phi(x_{u,i})}{\hat{p}_{u,i}(1-\hat{p}_{u,i})},$$
$$\mathcal{L}_{upo} \geq \mathcal{L}_{Bias} \quad \text{for } w_{u,i} = \frac{(e_{u,i}-\hat{e}_{u,i})}{\hat{p}_{u,i}},$$
$$\mathcal{L}_{upo} \geq \mathcal{L}_{Var\text{-}Tay} \quad \text{for } w_{u,i} = \frac{2(1-p_{u,i})(e_{u,i}-\hat{e}_{u,i})^2}{|\mathcal{D}|\cdot p_{u,i}^2}.$$

The $\mathcal{L}_{\text{BMSE-abs}}$ is defined below:

$$\mathcal{L}_{\text{BMSE-abs}} = \left|\frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}}\left(\frac{o_{u,i}}{\hat{p}_{u,i}} - \frac{1-o_{u,i}}{1-\hat{p}_{u,i}}\right)\phi(x_{u,i})\right|.$$

This Lemma shows that the $\mathcal{L}_{\text{UPO}}$ loss is able to capture heterogeneity of each metric and resolve metric optimization conflicts, forming the theoretical and methodological foundation for the unified minimax optimization.

Unified Minimax Propensity Optimization

We now propose our Unified Minimax Propensity Optimization (UPO) framework, designed to simultaneously optimize bias, variance, calibration, and balancing in propensity learning, i.e., addressing their optimization conflict, such as bias-variance trade-off.

Unified Minimax Optimization Formulation Our UPO-DR loss integrates standard cross-entropy (CE) with our unified bin-based residual loss:

$$\mathcal{L}_{\text{UPO}-\text{DR}}(\hat{p}, w) = \mathcal{L}_{\text{CE}}(\hat{p}, o) + \beta\cdot\mathcal{L}_{\text{UPO}}(\hat{p}, w), \quad (5)$$

where $\beta > 0$ balances predictive accuracy and robustness enforced by bin-based loss. Explicitly defined by Equation (4), the adaptive bin-specific weights $w_{u,i}$ encapsulate calibration, balancing, DR bias, and DR variance information. Specifically, we enforce robust optimization using a minimax objective:

$$\hat{p}_{\text{UPO}-\text{DR}} = \arg\min_{\hat{p}}\left\{\mathcal{L}_{\text{CE}}(\hat{p}, o) + \beta\max_{w}\mathcal{L}_{\text{UPO}}(\hat{p}, w)\right\}.$$

This formulation targets worst-case scenarios over adaptive weights $w_{u,i}$. By implicitly learning adversarial combinations of the four metrics within each bin, UPO ensures robust and balanced performance.

Unbiasedness and Variance Reduction We provide theoretical guarantees on unbiasedness and variance reduction properties of UPO-DR.

Theorem 1 (Unbiasedness of UPO Loss) *When the learned propensities are correct, as $|\mathcal{D}| \to \infty$ with finite bin number M, our unified bin-based loss converges to zero almost surely:*

$$\lim_{|\mathcal{D}|\to\infty}\mathcal{L}_{\text{UPO}}(\hat{p}, w) \to 0 \quad \textit{almost surely.}$$

Thus, with correctly learned propensities, $\mathcal{L}_{\text{UPO}}$ vanishes with increasing sample size. Next, we derive DR-UPO variance and its optimal variance value.

Theorem 2 (Variance Reduction of UPO-DR) *Given the UPO-DR loss as defined in Equation (5), its conditional variance is $\mathbb{V}(\mathcal{L}_{\text{UPO}-\text{DR}} \mid \mathbf{o}) = \mathbb{V}(\mathcal{L}_{\text{CE}}) + \beta^2\mathbb{V}(\mathcal{L}_{\text{UPO}}) + 2\beta\,\text{Cov}(\mathcal{L}_{\text{CE}}, \mathcal{L}_{\text{UPO}})$, optimizing this variance w.r.t. β yields:*

$$\beta^* = -\frac{\text{Cov}(\mathcal{L}_{\text{CE}}, \mathcal{L}_{\text{UPO}})}{\mathbb{V}(\mathcal{L}_{\text{UPO}})},$$

then with minimal achievable variance:

$$\mathbb{V}(\mathcal{L}_{\text{UPO}-\text{DR}} \mid \mathbf{o})|_{\beta=\beta^*} = (1-\rho^2)\mathbb{V}(\mathcal{L}_{\text{CE}}|\,\mathbf{o}),$$

smaller than $\mathbb{V}(\mathcal{L}_{\text{CE}}|\mathbf{o})$, with $\rho = Corr(\mathcal{L}_{\text{CE}}, \mathcal{L}_{\text{UPO}})$.

The above theorem rigorously demonstrates variance reduction compared to using base propensity training loss $\mathcal{L}_{\text{CE}}$ alone. Furthermore, we would like to compare the (potentially optimal) variance between our proposed UPO-DR loss and a naive way of simultaneously controlling the four metrics, i.e., simply adding them one by one after the basic cross entropy loss, named multi-metric loss defined below:

$$\mathcal{L}_{\text{multi-metric}} = \mathcal{L}_{\text{CE}} + \sum_{j=1}^{4}\lambda_j\frac{1}{|B_m|}\sum_{(u,i)\in B_m} w_{u,i}^{(j)}(o_{u,i}-\hat{p}_{u,i}),$$

where j represent our four metrics, Bias, Variance, ECE, and BMSE. In addition, $w_{u,i}^{j}$ is defined in Lemma 2.

Corollary 1 (Variance Comparison) *The optimal variance of UPO-DR is strictly smaller than that of the naive multi-metric estimator due to additional covariance terms between metrics: $\mathbb{V}(\mathcal{L}_{multi\text{-}metric})_{min} > \mathbb{V}(L_{\text{UPO}})_{min}$.*

Hence, our UPO-DR provides rigorous variance reduction relative to simpler multi-metric estimators.

Tightness Bounds

We rigorously quantify the approximation tightness between our proposed Unified Propensity Optimization (UPO) bin-based metrics and corresponding metrics (Bias, Variance, and BMSE-abs). We generally denote $\mathcal{L}_{\text{metric}}$ as the original metrics and denote $\mathcal{L}_{\text{UPO}}^{\text{metric}}$ as the loss in Equation 4 with corresponding weights in Lemma 2.

Theorem 3 (Tightness Bounds) *Given normalized feature balancing function, $\phi(x) \in [-1, 1]$, the bin-specific bounding factor is as follows:*

$$Q_m^{bmse} = \frac{1}{|\mathcal{B}_m|} \sum_{(u,i)\in\mathcal{B}_m} \frac{1}{\hat{p}_{u,i}(1-\hat{p}_{u,i})}, \quad \textit{for BMSE-abs},$$

$$Q_m^{bias} = \frac{1}{|\mathcal{B}_m|} \sum_{(u,i)\in\mathcal{B}_m} \frac{1}{\hat{p}_{u,i}}, \quad \textit{for Bias},$$

$$Q_m^{var} = \frac{2}{|\mathcal{D}|} \frac{1}{|\mathcal{B}_m|} \sum_{(u,i)\in\mathcal{B}_m} \frac{1}{\hat{p}_{u,i}}, \quad \textit{for Variance},$$

As $|\mathcal{D}| \to \infty$, with probability at least $1-\eta$, we have the following unified tightness bound for each metric:

$$\left|\mathcal{L}_{metric} - \mathcal{L}_{\text{UPO}}^{metric}\right| \leq \sqrt{\frac{2\sum_{m=1}^{M} |\mathcal{B}_m|^2 \left(Q_m^{metric}\right)^2 \log(2/\eta)}{|\mathcal{D}|^2}},$$

where metric $\in$ {Bias, Var, BMSE}. These explicit bounds rigorously validate that our UPO formulation effectively controls the original metrics.

Optimization Algorithm

This section formalizes the joint training pipeline of the UPO-DR framework, which coordinates four models: (1) adversarial weight model $w = \omega(x)$, (2) propensity model $\hat{p} = h_\psi(x)$, (3) prediction model $\hat{r} = f_\theta(x)$, and (4) imputation model $\hat{e} = g_\phi(x)$. The algorithm 1 alternates optimizing the adversarial weights and propensity model, then updates the prediction and imputation models.

Algorithm 1 Joint Training Algorithm for UPO-DR

Input:
- Set of user-item features X;
- Observation indicator matrix $\mathcal{O}$;
- All user-item pairs matrix $\mathcal{D}$;
- Observed outcomes for observed samples $\mathbf{R}^o$;
- Hyperparameters β;

while not converge **do**
- **for** number of training iterations **do**
  - Sample (u, i) pairs $\{(u_k, i_k)\}_{k=1}^{K}$ from $\mathcal{D}$;
  - update the adversarial weights w via maximizing $\mathcal{L}_{\text{UPO}}(\hat{p}, w)$;
  - update the propensity model h_ψ based on $\mathcal{L}_{\text{UPO}}(\hat{p}, w)$ with adversarial weights;
- **end for**

end while

while not converge **do**
- **for** number of training iterations **do**
  - Sample (u, i) pairs $\{(u_j, i_j)\}_{j=1}^{J}$ from $\mathcal{D}$;
  - Update the prediction model f_θ using DR loss, $\mathcal{L}_{\text{DR}}(\theta)$ in Equation 1;
  - Sample (u, i) pairs $\{(u_s, i_s)\}_{s=1}^{S}$ from $\mathcal{O}$;
  - Update the imputation model g_ϕ using imputation loss, $\mathcal{L}_{\text{imp}}(\phi)$ in Equation 2;
- **end for**

end while

Semi-synthetic Experiments

We conduct semi-synthetic experiments using the **MovieLens 100K (ML-100K)** dataset (Harper and Konstan 2015), consisting of 943 users and 1,682 movies, with 100,000 ratings between 1 and 5 being originally observed (missing rate 0.937). Our semi-synthetic study pursues two goals:

1. **Diagnose limitations of single-objective learning.** We demonstrate insufficiency for reliable propensity estimation if optimizing only one metric in bias, variance, calibration, and balancing, given MNAR feedback.
2. **Demonstrate the advantage of UPO.** We evaluate whether our UPO framework, delivers uniformly superior performance.

Experimental Setup. Following previous works (Schnabel et al. 2016; Guo et al. 2021), we first complete the full rating matrix R by Matrix Factorization (MF). Then we regard the completed synthetic rating matrix serves as the ground truth rating matrix R_{gt}. Then we map each rating to a 5-scale, based on the original rating quantile in R_{gt}, to align the rating distribution with the unbiased data in **Yahoo! R3** dataset. To generate the ground truth propensity scores $p_{u,i}$, we set $p_{u,i} = p^{k(r_{u,i})}$, where $k(r_{u,i}) = 1, 2, 3, 4, 4$ for rating $r_{u,i} = 5, 4, 3, 2, 1$. In our experiment, $p \in \{0.4, 0.5, 0.6\}$, representing varying levels of selection bias. We resample observations from the assigned propensity scores to conduct the observation indicator, i.e., $o_{u,i} \sim \text{Bernoulli}(p_{u,i})$ to obtain the observed data.

Experimental Details and Baselines. We evaluate the performance of our method along with other propensity-based methods controlling one of the four metrics. Specifically, we compare the following six methods: DR-JL, the baseline method with no additional control; DR-V2, optimized for balancing; DCE-DR, optimized for calibration; DR-BIAS, optimized for bias; MRDR, optimized for variance; UPO (Ours), controlling all four metrics.

Performance Analysis. Figure 1 and Table 1 present the evaluation results of these six estimators under different propensity settings. As demonstrated, single-objective methods such as DCE-DR achieve excellent calibration, but exhibit poor performance in the other 3 metrics when $p = 0.4$. Similar trends appear across other estimators and metrics. In contrast, UPO well addresses the conflict between the four metrics, achieving acceptable performance on all metrics, and thus highlights the robustness of the UPO under data MNAR. The results for $p = 0.5, 0.6$ can be found in our arXiv version.

	$p = 0.4$			
Method	ECE	BMSE	Bias	Var
DCE-DR	**0.2207**	6.5044	0.0289	2.1499e-4
DR-BIAS	0.2249	5.6342	**0.0216**	1.9783e-4
DR-JL	0.2490	6.6144	0.0642	4.6684e-4
DR-V2	0.2323	**0.6419**	0.1857	1.6180e-4
MRDR	0.2271	6.4284	0.0357	<u>1.0205e-4</u>
UPO (Ours)	<u>0.2232</u>	<u>5.5740</u>	<u>0.0238</u>	**0.5169e-4**

Table 1: Semi-synthetic experiment with $p = 0.4$. The best ones are in bold and the second best are underlined.

Real-world Experiments

Experimental Settings

Datasets We conduct experiments on three benchmark datasets widely used in debiased recommendation tasks: **Coat** (Steck 2010), **Yahoo! R3** (Schnabel et al. 2016) , including both MNAR and unbiased (MAR) ratings. Specifically, **Coat** includes 290 users and 300 items with 6,960 MNAR and 4,640 MAR ratings, and **Yahoo! R3** contains 5,400 users and 1,000 items with 311,704 MNAR and 54,000 MAR ratings. We binarize the ratings in **Coat** and **Yahoo! R3** datasets ≥ 3 to 1, and others to 0. Additionally, we have also conduct our methods on the **KuaiRec** (Gao et al. 2022) dataset to further demonstrate the generality of our approach, including 1,411 users and 3,327 items, which consists of a biased subset with 201,171 records and an unbiased subset with 117,113 records. We binarize the ratings in **KuaiRec** datasets greater than two to 1, and others to 0. The results of **Kuairec** can be found in the arXiv version.

Baselines We compare our method against the following representative baseline methods building on MF backbone (Koren, Bell, and Volinsky 2009): IPS (Schnabel et al. 2016), SNIPS (Swaminathan and Joachims 2015), ASIPS (Saito 2020a), IPS-V2 (Li et al. 2023d), KBIPS, AKBIPS (Li et al. 2024e), DR (Saito 2020b), DR-JL (Wang et al. 2019b), MRDR-JL (Guo et al. 2021), DR-BIAS and DR-MSE (Dai et al. 2022), MR (Li et al. 2023b), TDR (Li et al. 2023c), TDR-JL (Li et al. 2023c), StableDR (Li, Zheng, and Wu 2023), DR-V2 (Li et al. 2024a), KBDR and AKBDR (Li et al. 2024e), DCE-DR and DCE-TDR (Kweon and Yu 2024), and Cali-MR (Gong and Ma 2025).

Evaluation Metrics We adopt three standard evaluation metrics for recommendation performance: **AUC**, **NDCG@T**, **F1@T**. We set $T = 5$ for **Coat** and **Yahoo! R3** datasets, and $T = 20$ for **KuaiRec** dataset.

Implementation Details We use the same hyperparameter search space and follow the results in Cali-MR (Gong and Ma 2025) and we optimize penalization hyperparameters by Optuna (Akiba et al. 2019). We adopt an equal-width binning strategy. The adversarial weights are trained with L2 regularization. The learning rate is selected from $\{0.01, 0.05\}$, and weight decaying is tuned within $\{1 \times 10^{-6}, 5 \times 10^{-6}, 1 \times 10^{-5}, \ldots, 1 \times 10^{-3}, 5 \times 10^{-3}\}$. Experiments were performed on NVIDIA A100 GPUs.

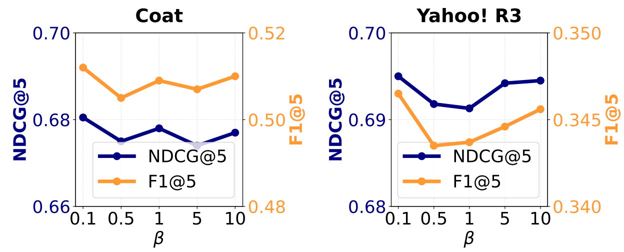


Figure 2: Parameter sensitivity analysis on β.

Performance Analysis

Table 2 compares our proposed UPO method with existing state-of-the-art estimators across two datasets. Clearly, all IPS- and DR-based methods outperform the naive baseline, showing the necessity of addressing MNAR bias. Among these, DR-BIAS, MRDR, and DR-MSE further refine estimation by controlling bias, variance, and their trade-off. DCE-TDR introduces propensity calibration into targeted learning, while AKBDR employs adaptive kernel balancing to enhance balancing.

Nevertheless, our UPO estimator consistently achieves superior performance across all cases. Unlike methods addressing only one or two criteria, UPO explicitly manages the complex trade-offs among calibration, balancing, DR bias, and DR variance using a unified minimax approach. This multi-criteria optimization leads to more robust and accurate propensity estimates, demonstrating clear advantages over existing state-of-the-art approaches.

Time and Space Analysis

We investigate the training time (in seconds) and the number of learnable parameters of the compared methods. Experiments were conducted on NVIDIA A100 GPUs, with results shown in Table 3. We observe that our proposed UPO estimator does not significantly increase computational complexity compared to other methods. Specifically, UPO maintains a comparable computational time vs other debiasing estimators, like DR-V2 and DCE-DR, across three datasets. UPO offers superior performance with minimal computational overhead.

Parameter Sensitivity Analysis

We investigate the sensitivity of the UPO framework to its key regularization hyperparameter β in Equation (4). Figure 2 illustrates the changes in model performance on the **Coat** and **Yahoo! R3** datasets, as β varies in $\{0.1, 0.5, 1, 5, 10\}$. On Coat and Yahoo! R3 datasets, despite changes in β, variations in **NDCG@5** and **F1@5** are minor, demonstrating strong robustness to the choice of β. The sensitivity analysis shows that our model is robust and reliable in applications.

Effect of the Number of Bins (M)

Figure 3 plots how four metrics and AUC are affected by varying numbers of spitted bin on **Yahoo! R3** dataset. As M increasing, Figure 3 shows clear reductions in ECE, BMSE,

Method	Coat AUC	Coat NDCG@5	Coat F1@5	Yahoo! R3 AUC	Yahoo! R3 NDCG@5	Yahoo! R3 F1@5
Naive	0.703±0.006	0.605±0.012	0.467±0.007	0.673±0.001	0.635±0.002	0.306±0.002
IPS	0.717±0.007	0.617±0.009	0.473±0.008	0.678±0.001	0.638±0.002	0.318±0.002
SNIPS	0.714±0.012	0.614±0.012	0.474±0.009	0.683±0.002	0.639±0.002	0.316±0.002
ASIPS	0.719±0.009	0.618±0.012	0.476±0.009	0.679±0.003	0.640±0.003	0.319±0.003
IPS-V2	0.726±0.005	0.627±0.009	0.479±0.008	0.685±0.002	0.646±0.003	0.320±0.002
KBIPS	0.714±0.003	0.618±0.010	0.474±0.007	0.676±0.002	0.642±0.003	0.318±0.002
AKBIPS	0.732±0.004	0.636±0.006	0.483±0.006	0.689±0.001	0.658±0.002	0.324±0.002
DR	0.718±0.008	0.623±0.009	0.474±0.007	0.684±0.002	0.658±0.003	0.326±0.002
DR-JL	0.723±0.005	0.629±0.007	0.479±0.005	0.685±0.002	0.653±0.002	0.324±0.002
MRDR-JL	0.727±0.005	0.627±0.008	0.480±0.008	0.684±0.002	0.652±0.003	0.325±0.002
DR-BIAS	0.726±0.004	0.629±0.009	0.482±0.007	0.685±0.002	0.653±0.002	0.325±0.003
DR-MSE	0.727±0.007	0.631±0.008	0.484±0.007	0.687±0.002	0.657±0.003	0.327±0.003
MR	0.724±0.004	0.636±0.006	0.481±0.006	0.691±0.002	0.647±0.002	0.316±0.003
TDR	0.714±0.006	0.634±0.011	0.483±0.008	0.688±0.003	0.662±0.002	0.329±0.002
TDR-JL	0.731±0.005	0.639±0.007	0.484±0.007	0.689±0.002	0.656±0.004	0.327±0.003
StableDR	0.735±0.005	0.640±0.007	0.484±0.006	0.688±0.002	0.661±0.003	0.329±0.002
DR-V2	0.734±0.007	0.639±0.009	0.487±0.006	0.690±0.002	0.660±0.005	0.328±0.002
KBDR	0.730±0.003	0.631±0.005	0.482±0.006	0.682±0.002	0.648±0.003	0.323±0.002
AKBDR	0.745±0.004	0.645±0.008	0.493±0.007	0.692±0.002	0.661±0.002	0.328±0.002
DCE-DR	0.736±0.006	0.648±0.007	0.489±0.005	0.698±0.002	0.670±0.002	0.333±0.003
DCE-TDR	0.740±0.004	0.651±0.006	0.489±0.007	0.701±0.002	0.672±0.002	0.331±0.002
Cali-MR	0.741±0.002	0.658±0.004	0.495±0.004	0.703±0.002	0.678±0.002	0.338±0.004
UPO (Ours)	**0.749±0.003**	**0.691±0.002**	**0.515±0.002**	**0.717±0.003**	**0.694±0.004**	**0.345±0.003**

Table 2: Performance comparison on **Coat** and **Yahoo! R3** on AUC, NDCG@K and F1@K. The best results are in bold. The standard deviation is obtained from 10 repeated experiments.

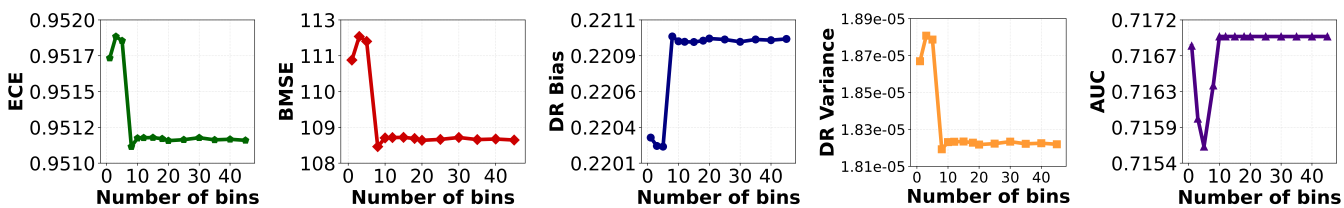


Figure 3: Sensitivity of UPO to the number of bins M on **Yahoo! R3**. We report five evaluation criteria, ECE, BMSE, bias, variance, and AUC, as M increases from 1 to 45 on **Yahoo! R3** dataset. Performance stabilises once $M \geq 10$.

	Coat Time	Coat Params	Yahoo! R3 Time	Yahoo! R3 Params	KuaiRec Time	KuaiRec Params
MRDR_JL	1.82	56K	445.3	1574K	452.4	1704K
DCE-DR	5.71	57K	590.6	787K	795.0	3411K
DR-BIAS	1.80	56K	580.7	1574K	901.5	3409K
DR-V2	6.69	56K	500.5	4723K	902.1	9092K
DR-JL	4.52	28K	493.0	1574K	843.1	3409K
UPO (Ours)	5.02	113K	520.9	3149K	580.2	6005K

Table 3: Time and space analysis. Time (in s) is used for the training duration, and #params denotes the number of learnable parameters.

and variance, and a rise in AUC. Interestingly, an increase in bias suggests we only need to split a few bins, which is also a trade-off in practice between the performance of bias and other metrics. In addition, the performance will become stable when the number of bins exceeds a threshold.

Conclusion

In this paper, we propose a Unified Propensity Optimization (UPO) framework to improve the accuracy of propensity score prediction. Recognizing limitations of existing methods optimizing isolated metrics, we propose a unified formulation integrating four critical criteria: ECE (calibration), BMSE (feature balancing), bias, and variance. We revealed structural commonalities among these criteria, enabling unified adaptive weighting. Then we developed a minimax optimization approach with adversarial weighting. The theoretical guarantee of our method shows that UPO can control bias and reduce variance compared to naive multi-metric combinations. Experiments on semi-synthetic and real-world datasets, including an industrial dataset, validated UPO's superior empirical performance over several SOTA propensity-based methods. One potential limitation is that we can adopt a dynamic bin-splitting strategy to adapt our minimax algorithm, rather than pre-specified bins.

References

Ai, Q.; Bi, K.; Luo, C.; Guo, J.; and Croft, W. B. 2018. Unbiased Learning to Rank with Unbiased Propensity Estimation. In *SIGIR*.

Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; and Koyama, M. 2019. Optuna: A next-generation hyperparameter optimization framework. In *KDD*.

Bonner, S.; and Vasile, F. 2018. Causal embeddings for recommendation. In *RecSys*.

Chen, J.; Dong, H.; Wang, X.; et al. 2022. Bias and Debias in Recommender System: A Survey and Future Directions. *ACM Transactions on Information Systems*.

Dai, Q.; Li, H.; Wu, P.; et al. 2022. A Generalized Doubly Robust Learning Framework for Debiasing Post-click Conversion Rate Prediction. In *KDD*.

Ding, S.; Wu, P.; Feng, F.; He, X.; Wang, Y.; Liao, Y.; and Zhang, Y. 2022. Addressing Unmeasured Confounder for Recommendation with Sensitivity Analysis. In *KDD*.

Gao, C.; Li, S.; Lei, W.; Chen, J.; Li, B.; Jiang, P.; He, X.; Mao, J.; and Chua, T.-S. 2022. KuaiRec: A fully-observed dataset and insights for evaluating recommender systems. In *CIKM*.

Gong, S.; and Ma, C. 2025. Gradient-Based Multiple Robust Learning Calibration on Data Missing-Not-at-Random via Bi-Level Optimization. *Entropy*.

Guo, C.; Pleiss, G.; Sun, Y.; and Weinberger, K. Q. 2017. On calibration of modern neural networks. In *ICML*.

Guo, S.; Zou, L.; Liu, Y.; et al. 2021. Enhanced Doubly Robust Learning for Debiasing Post-click Conversion Rate Estimation. In *SIGIR*.

Harper, F. M.; and Konstan, J. A. 2015. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*.

Hu, W.; Sun, X.; Liu, Q.; Wu, L.; and Wang, L. 2025. Uncertainty calibration for counterfactual propensity estimation in recommendation. *IEEE Transactions on Knowledge and Data Engineering*.

Huang, S.; Li, H.; Li, Q.; Zheng, C.; and Liu, L. 2023. Pareto invariant representation learning for multimedia recommendation. In *ACM MM*.

Huang, S.; Li, H.; Zheng, C.; Ge, M.; Gao, W.; Wang, L.; and Liu, L. 2025a. Text-Driven Fashion Image Editing with Compositional Concept Learning and Counterfactual Abduction. In *CVPR*.

Huang, S.; Li, H.; Zheng, C.; Wang, L.; Liao, G.; Gong, Z.; Yang, H.; and Liu, L. 2025b. Visual Representation Learning through Causal Intervention for Controllable Image Editing. In *CVPR*.

Imai, K.; and Ratkovic, M. 2014. Covariate balancing propensity score. *Journal of the Royal Statistical Society Series B: Statistical Methodology*.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer*.

Kweon, W.; and Yu, H. 2022. Obtaining calibrated probabilities with personalized ranking models. In *AAAI*.

Kweon, W.; and Yu, H. 2024. Doubly Calibrated Estimator for Recommendation on Data Missing Not at Random. In *WWW*.

Li, H.; Dai, Q.; Li, Y.; Lyu, Y.; Dong, Z.; Zhou, X.-H.; and Wu, P. 2023a. Multiple Robust Learning for Recommendation. In *AAAI*.

Li, H.; Dai, Q.; Li, Y.; et al. 2023b. Multiple Robust Learning for Recommendation. In *AAAI*.

Li, H.; Lyu, Y.; Zheng, C.; and Wu, P. 2023c. TDR-CL: Targeted Doubly Robust Collaborative Learning for Debiased Recommendations. In *ICLR*.

Li, H.; Xiao, Y.; Zheng, C.; and Wu, P. 2024a. Relaxing Accurate Imputation Assumption in Doubly Robust Learning. In *ICML*.

Li, H.; Xiao, Y.; Zheng, C.; Wu, P.; and Cui, P. 2023d. Propensity Matters: Measuring and Enhancing Balancing for Recommendation. In *ICML*.

Li, H.; Zheng, C.; Ding, S.; Feng, F.; He, X.; Geng, Z.; and Wu, P. 2024b. Be Aware of the Neighborhood Effect: Modeling Selection Bias under Interference for Recommendation. In *ICLR*.

Li, H.; Zheng, C.; Wang, S.; Wu, K.; Wang, E.; Wu, P.; Geng, Z.; Chen, X.; and Zhou, X.-H. 2024c. Relaxing the accurate imputation assumption in doubly robust learning for debiased collaborative filtering. In *ICML*.

Li, H.; Zheng, C.; Wang, W.; Wang, H.; Feng, F.; and Zhou, X.-H. 2024d. Debiased recommendation with noisy feedback. In *KDD*.

Li, H.; Zheng, C.; and Wu, P. 2023. StableDR: Stabilized Doubly Robust Learning for Recommendation. In *ICLR*.

Li, H.; Zheng, C.; Wu, P.; Kuang, K.; Liu, Y.; and Cui, P. 2023e. Who should be Given Incentives? Counterfactual Optimal Treatment Regimes Learning for Recommendation. In *KDD*.

Li, H.; Zheng, C.; Xiao, Y.; and Wu, P. 2024e. Debiased Collaborative Filtering with Kernel-based Causal Balancing. In *ICLR*.

Lin, J.; Dai, X.; Shan, R.; Chen, B.; Tang, R.; Yu, Y.; and Zhang, W. 2025a. Large language models make sample-efficient recommender systems. *Frontiers of Computer Science*.

Lin, J.; Dai, X.; Xi, Y.; Liu, W.; Chen, B.; Zhang, H.; Liu, Y.; Wu, C.; Li, X.; Zhu, C.; et al. 2025b. How can recommender systems benefit from large language models: A survey. *ACM Transactions on Information Systems*.

Ma, X.; Zhao, L.; Huang, G.; Wang, Z.; Hu, Z.; Zhu, X.; and Gai, K. 2018. Entire Space Multi-Task Model: An Effective Approach for Estimating Post-Click Conversion Rate. In *SIGIR*.

Naeini, M. P.; Cooper, G.; and Hauskrecht, M. 2015. Obtaining well calibrated probabilities using bayesian binning. In *AAAI*.

Pan, H.; Zheng, C.; Wang, W.; Jiang, J.; Li, X.; Li, H.; and Feng, F. 2025. Batch-Adaptive Doubly Robust Learning for Debiasing Post-Click Conversion Rate Prediction Under Sparse Data. *ACM Transactions on Information Systems*.

Saito, Y. 2020a. Asymmetric Tri-training for Debiasing Missing-Not-At-Random Explicit Feedback. In *SIGIR*.

Saito, Y. 2020b. Doubly Robust Estimator for Ranking Metrics with Post-Click Conversions. In *RecSys*.

Saito, Y.; Yaginuma, S.; Nishino, Y.; Sakata, H.; and Nakata, K. 2020. Unbiased Recommender Learning from Missing-Not-at-Random Implicit Feedback. In *WSDM*.

Schnabel, T.; Swaminathan, A.; Singh, A.; Chandak, N.; and Joachims, T. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *ICML*.

Steck, H. 2010. Training and Testing of Recommender Systems on Data Missing Not at Random. In *KDD*.

Su, J.; Chen, C.; Lin, Z.; Li, X.; Liu, W.; and Zheng, X. 2023. Personalized behavior-aware transformer for multi-behavior sequential recommendation. In *ACM MM*.

Swaminathan, A.; and Joachims, T. 2015. The Self-Normalized Estimator for Counterfactual Learning. In *NeurIPS*.

Wang, H. 2024. Improving Neural Network Generalization on Data-Limited Regression with Doubly-Robust Boosting. In *AAAI*.

Wang, H.; Chang, T.-W.; Liu, T.; Huang, J.; Chen, Z.; Yu, C.; Li, R.; and Chu, W. 2022. ESCM2: Entire Space Counterfactual Multi-task Model for Post-Click Conversion Rate Estimation. In *SIGIR*.

Wang, H.; Chen, Z.; Liu, Z.; Chen, X.; Li, H.; and Lin, Z. 2025a. Proximity matters: Local proximity enhanced balancing for treatment effect estimation. In *KDD*.

Wang, H.; Chen, Z.; Liu, Z.; Li, H.; Yang, D.; Liu, X.; and Li, H. 2024. Entire Space Counterfactual Learning for Reliable Content Recommendations. *IEEE Transactions on Information Forensics and Security*.

Wang, H.; Chen, Z.; Wang, H.; Tan, Y.; Pan, L.; Liu, T.; Chen, X.; Li, H.; and Lin, Z. 2025b. Unbiased Recommender Learning from Implicit Feedback via Weakly Supervised Learning. In *ICML*.

Wang, H.; Chen, Z.; Zhang, H.; Li, Z.; Pan, L.; Li, H.; and Gong, M. 2025c. Debiased Recommendation via Wasserstein Causal Balancing. *ACM Transactions on Information Systems*.

Wang, X.; Zhang, R.; Sun, Y.; and Qi, J. 2019a. Doubly Robust Joint Learning. In *ICML*.

Wang, X.; Zhang, R.; Sun, Y.; and Qi, J. 2019b. Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random. In *ICML*.

Wang, Z.; Chen, X.; Wen, R.; Huang, S.-L.; Kuruoglu, E. E.; and Zheng, Y. 2020. Information theoretic counterfactual learning from missing-not-at-random feedback. In *NeurIPS*.

Wu, P.; Li, H.; Zheng, C.; Zeng, Y.; Chen, J.; Liu, Y.; Guo, R.; and Zhang, K. 2025. Learning Counterfactual Outcomes Under Rank Preservation. In *NeurIPS*.

Yang, H.; Zheng, C.; Liao, G.; Huang, S.; Liao, J.; Gong, Z.; Li, H.; and Liu, L. 2025. CAP: Causal Air Quality Index Prediction Under Interference with Unmeasured Confounding. In *WWW*.

Yang, L.; Cui, Y.; Xuan, Y.; Wang, C.; Belongie, S.; and Estrin, D. 2018. Unbiased offline recommender evaluation for missing-not-at-random implicit feedback. In *RecSys*.

Zhang, H.; Wang, S.; Li, H.; Zheng, C.; Chen, X.; Liu, L.; Luo, S.; and Wu, P. 2024. Uncovering the Propensity Identification Problem in Debiased Recommendations. In *ICDE*.

Zhang, S.; and Xia, T. 2025. CBPL: A Unified Calibration and Balancing Propensity Learning Framework in Causal Recommendation for Debiasing. In *IJCAI Workshop on Causal Learning RecSys*.

Zhang, S.; Zhang, Y.; Chen, J.; and Sui, H. 2025. Addressing Correlated Latent Exogenous Variables in Debiased Recommender Systems. In *KDD*.

Zheng, C.; Hu, T.; Zhou, C.; and Zhou, X.-H. 2024. UMVUE-DR: Uniformly Minimum Variance Unbiased Doubly Robust Learning with Reduced Bias and Variance for Debiased Recommendation. In *ICDM Workshop on Causal Representation Learning*.

Zheng, C.; Pan, H.; Zhang, Y.; and Li, H. 2025a. Adaptive Structure Learning with Partial Parameter Sharing for Post-Click Conversion Rate Prediction. In *SIGIR*.

Zheng, C.; Yang, H.; Li, H.; and Yang, M. 2025b. Unveiling Extraneous Sampling Bias with Data Missing-Not-At-Random. In *NeurIPS*.

Zheng, X.; Su, J.; Liu, W.; and Chen, C. 2022. DDGHM: Dual dynamic graph with hybrid metric training for cross-domain sequential recommendation. In *ACM MM*.

Zhou, C.; Li, H.; Yao, L.; and Gong, M. 2025a. Counterfactual Implicit Feedback Modeling. In *NeurIPS*.

Zhou, C.; Li, Y.; Zheng, C.; Zhang, H.; Zhang, M.; Li, H.; and Gong, M. 2025b. A Two-Stage Pretraining-Finetuning Framework for Treatment Effect Estimation with Unmeasured Confounding. In *KDD*.

Zhou, Y.; Feng, T.; Liu, M.; and Zhu, Z. 2023. A Generalized Propensity Learning Framework for Unbiased Post-Click Conversion Rate Estimation. In *CIKM*.